Assignment4

November 13, 2025

0.0.1 Question 1 (a)

[81]:
```python
import pandas as pd
import numpy as np
```

[83]:
```python
COMMODITIES = {
    'Oil': 'Oil.csv',
    'Wheat': 'Wheat.csv',
    'NaturalGas': 'NaturalGas.csv',
    'Copper': 'Copper.csv',
    'Silver': 'Silver.csv'
}

PRICE_COLUMN = 'Adj Close'

START_DATE = '2019-01-02'
END_DATE = '2025-10-31'

# an index with every single calendar day
all_calendar_days = pd.date_range(start=START_DATE, end=END_DATE, freq='D')

# empty DataFrame to hold our aligned price data
aligned_prices = pd.DataFrame(index=all_calendar_days)
```

[85]:
```python
# Load, align, and fill each commodity
for name, filename in COMMODITIES.items():

    df = pd.read_csv(filename)

    # 'Date' column to datetime objects
    df['Date'] = pd.to_datetime(df['Date'])

    # Set the 'Date' as the index
    df = df.set_index('Date')

    df_renamed = df[[PRICE_COLUMN]].rename(columns={PRICE_COLUMN: name})
    df_aligned = df_renamed.reindex(all_calendar_days)

```

```
    # Fill missing values (NaNs) by carrying forward the last available price.
    df_filled = df_aligned.ffill()

    # Add this commodity's filled data to our main DataFrame
    aligned_prices = aligned_prices.join(df_filled)

for col in aligned_prices.columns:
    aligned_prices[col] = aligned_prices[col].astype(str).str.replace(r'[$,]',␣
↪'', regex=True)
    aligned_prices[col] = pd.to_numeric(aligned_prices[col], errors='coerce')

aligned_prices = aligned_prices.bfill()

print("--- Aligned Daily Prices (Head) ---")
print(aligned_prices.head())
print("\n--- Aligned Daily Prices (Tail) ---")
print(aligned_prices.tail())
```

```
--- Aligned Daily Prices (Head) ---
                  Oil   Wheat  NaturalGas  Copper  Silver
2019-01-02  46.540001  506.75       2.958  2.6250  15.542
2019-01-03  47.090000  513.75       2.945  2.5705  15.706
2019-01-04  47.959999  517.00       3.044  2.6515  15.695
2019-01-05  47.959999  517.00       3.044  2.6515  15.695
2019-01-06  47.959999  517.00       3.044  2.6515  15.695

--- Aligned Daily Prices (Tail) ---
                  Oil   Wheat  NaturalGas  Copper     Silver
2025-10-27  61.310001  526.00       3.442  5.1405  46.562000
2025-10-28  60.150002  529.00       3.345  5.1405  47.125000
2025-10-29  60.480000  532.25       3.815  5.2335  47.721001
2025-10-30  60.570000  524.25       3.956  5.0780  48.428001
2025-10-31  60.980000  534.00       4.124  5.0655  47.993999
```

[87]:

```
# daily returns
daily_returns = aligned_prices.pct_change()
daily_returns = daily_returns.dropna(how='all')

print("\n--- Daily Returns (Head) ---")
print(daily_returns.head())
```

```
--- Daily Returns (Head) ---
                 Oil     Wheat  NaturalGas    Copper    Silver
2019-01-03  0.011818  0.013814   -0.004395 -0.020762  0.010552
2019-01-04  0.018475  0.006326    0.033616  0.031511 -0.000700
2019-01-05  0.000000  0.000000    0.000000  0.000000  0.000000
2019-01-06  0.000000  0.000000    0.000000  0.000000  0.000000
```

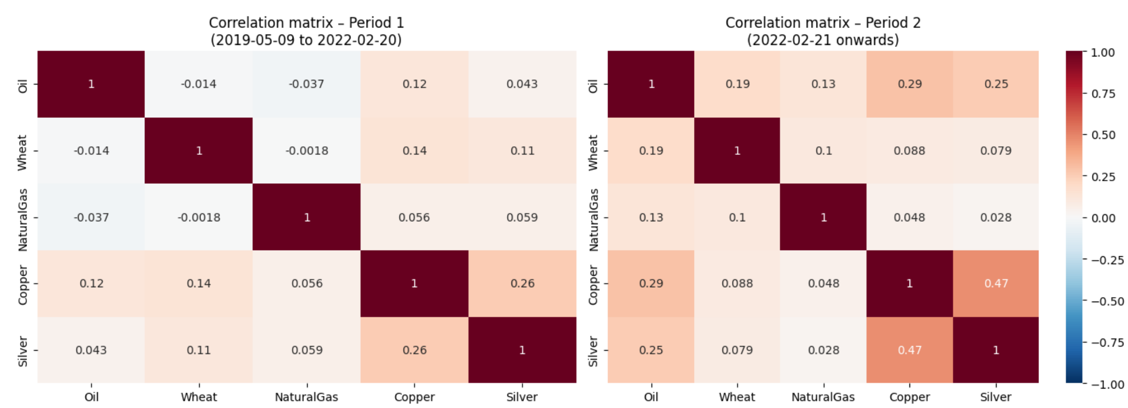
```
2019-01-07  0.011676 -0.000484    0.000329 -0.003960 -0.001657
```

[89]:
```python
# periods and data split
PERIOD_1_START = '2019-05-09'
PERIOD_1_END = '2022-02-20'
PERIOD_2_START = '2022-02-21'

# Slice the returns DataFrame using the period dates
returns_p1 = daily_returns.loc[PERIOD_1_START:PERIOD_1_END]
returns_p2 = daily_returns.loc[PERIOD_2_START:]

# Correlation matrices
# a matrix of the correlations from period 1
corr_p1 = returns_p1.corr()

print(f"\n--- (a.i) Correlation Matrix: Period 1 (ends {PERIOD_1_END}) ---")
print(corr_p1)


# a matrix of the correlations from period 2
corr_p2 = returns_p2.corr()

print(f"\n--- (a.ii) Correlation Matrix: Period 2 (starts {PERIOD_2_START}) ␣
↪---")
print(corr_p2)
```

```
--- (a.i) Correlation Matrix: Period 1 (ends 2022-02-20) ---
                 Oil     Wheat  NaturalGas    Copper    Silver
Oil         1.000000 -0.014220   -0.036527  0.117483  0.042781
Wheat      -0.014220  1.000000   -0.001834  0.139621  0.109378
NaturalGas -0.036527 -0.001834    1.000000  0.056159  0.059356
Copper      0.117483  0.139621    0.056159  1.000000  0.256016
Silver      0.042781  0.109378    0.059356  0.256016  1.000000

--- (a.ii) Correlation Matrix: Period 2 (starts 2022-02-21) ---
                 Oil     Wheat  NaturalGas    Copper    Silver
Oil         1.000000  0.185072    0.130849  0.293544  0.254489
Wheat       0.185072  1.000000    0.103932  0.088163  0.078951
NaturalGas  0.130849  0.103932    1.000000  0.047771  0.027610
Copper      0.293544  0.088163    0.047771  1.000000  0.466376
Silver      0.254489  0.078951    0.027610  0.466376  1.000000
```

[91]:
```python
import matplotlib.pyplot as plt
import seaborn as sns


fig, axes = plt.subplots(1, 2, figsize=(14, 5))
```

```python
# Period 1 heatmap
sns.heatmap(
    corr_p1,
    ax=axes[0],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
    annot=True,
    cbar=False
)
axes[0].set_title(f"Correlation matrix - Period 1\n({PERIOD_1_START} to {PERIOD_1_END})")

# Period 2 heatmap
sns.heatmap(
    corr_p2,
    ax=axes[1],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
    annot=True,
    cbar=True
)
axes[1].set_title(f"Correlation matrix - Period 2\n({PERIOD_2_START} onwards)")

plt.tight_layout()
plt.show()
```



[54]:
```python
corr_diff = corr_p2 - corr_p1

print("\n--- (a.iii) Change in Correlation (Period 2 - Period 1) ---")
print(corr_diff)
```

```
--- (a.iii) Change in Correlation (Period 2 - Period 1) ---
                 Oil     Wheat  NaturalGas    Copper    Silver
Oil         0.000000  0.199292    0.167376  0.176062  0.211708
Wheat       0.199292  0.000000    0.105766 -0.051458 -0.030427
NaturalGas  0.167376  0.105766    0.000000 -0.008387 -0.031745
Copper      0.176062 -0.051458   -0.008387  0.000000  0.210359
Silver      0.211708 -0.030427   -0.031745  0.210359  0.000000
```

Yes, the correlation matrix changed significantly between Period 1 (pre-invasion) and Period 2 (post-invasion). The main observation is an increase in the correlations between energy commodities (Oil, Natural Gas) and wheat, and a general increase of oil's correlation with all other assets.

By comparing the two matrices, we can quantify these changes: - Oil-Wheat: The correlation flipped from near-zero and slightly negative (-0.015) to significantly positive (0.185). - Oil-Natural Gas: This pair also flipped from slightly negative (-0.033) to a notable positive correlation (0.131). - Natural Gas-Wheat: This relationship, a key economic linkage, improved from non-existent to positive (0.104). - Oil and Metals: Oil's correlation with both metals increased substantially. Oil-Copper more than doubled from 0.115 to 0.294, and Oil-Silver saw a massive relative jump from 0.042 to 0.254. - Inter-Metal: The correlation between Copper and Silver, which was already positive, strengthened considerably from 0.262 to 0.466.

Reasoning: The Russian invasion of Ukraine in February 2022 was a major geopolitical and supply-side shock to the global economy, and its effects explain these new correlation structures.

- Russia is a top global exporter of crude oil, natural gas, and wheat. Ukraine is also a critical exporter of wheat. The war and the subsequent sanctions created simultaneous, massive supply disruptions and price uncertainty for all three of these commodities. Markets began to trade them in lockstep, as news affecting the war would impact all of them. This directly explains why their mutual correlations (Oil-Wheat, Oil-NG, NG-Wheat) all surged from near-zero into positive territory.
- The shock amplified existing, but previously dormant, economic links. The most prominent example is the Natural Gas-Wheat relationship. Natural gas is the primary input for producing nitrogen-based fertilizers. As the price of natural gas skyrocketed due to the European energy crisis, the cost of fertilizer soared, directly pushing up the production cost and, thus, the price of wheat. This new, tight cost-linkage is reflected in the correlation moving from -0.0018 to 0.104.
- The energy and food price shocks triggered a global spike in inflation. In this high-inflation environment, investors often flee to real assets (commodities) as an inflation hedge.

Oil, as the primary driver of this inflation shock and a barometer for geopolitical risk, became the central asset. This explains why its correlation with everything else, including the inflation hedge metals (Copper and Silver), increased so dramatically. Also, the strengthening correlation between Copper and Silver (0.262 to 0.466) also supports this. Investors began to group all real assets together, and when inflation fears rose (often led by oil), money flowed into the entire commodity complex, causing their prices to move more closely together. In summary, the correlation structure was fundamentally altered by the war, which tightly bound the energy and agricultural markets through a common supply shock and linked the entire commodity complex together as a single inflation trade.

[ ]:

0.0.2 Question 1 (b)

[59]:
```
SP500_FILENAME = 'SP500.csv'
PRICE_COLUMN = 'Adj Close'

# Load and process the S&P 500 data
df_sp = pd.read_csv(SP500_FILENAME)
df_sp['Date'] = pd.to_datetime(df_sp['Date'])

# same robust cleaning step we used before
df_sp[PRICE_COLUMN] = df_sp[PRICE_COLUMN].astype(str).str.replace(r'[$,]', '',␣
↪regex=True)
df_sp[PRICE_COLUMN] = pd.to_numeric(df_sp[PRICE_COLUMN], errors='coerce')

# Align to the master calendar
df_sp = df_sp.set_index('Date')

# Rename the price column
df_sp_renamed = df_sp[[PRICE_COLUMN]].rename(columns={PRICE_COLUMN: 'SP500'})
df_sp_aligned = df_sp_renamed.reindex(all_calendar_days)

# Fill missing values
df_sp_filled = df_sp_aligned.ffill()

# Backfill any NaNs at the very start
df_sp_filled = df_sp_filled.bfill()

print("--- S&P 500 Aligned Prices (Head) ---")
print(df_sp_filled.head())
```

```
--- S&P 500 Aligned Prices (Head) ---
                  SP500
2019-01-02  2510.030029
2019-01-03  2447.889893
2019-01-04  2531.939941
2019-01-05  2531.939941
2019-01-06  2531.939941
```

[61]:
```
# Calculate S&P 500 daily returns
sp500_returns = df_sp_filled.pct_change()
combined_returns = daily_returns.join(sp500_returns)

# Drop the first row which is all NaN
combined_returns = combined_returns.dropna(how='all')

print("\n--- Combined Returns with S&P 500 (Head) ---")
print(combined_returns.head())
```

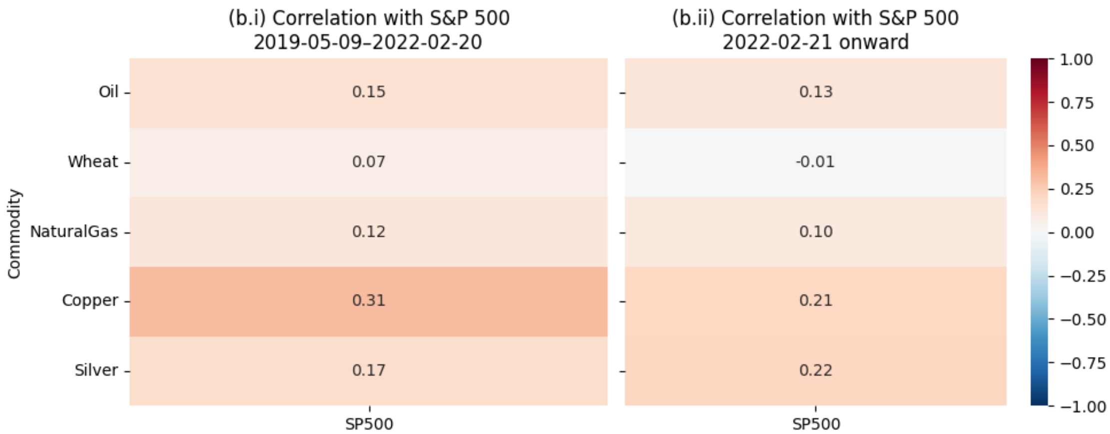
```
--- Combined Returns with S&P 500 (Head) ---
                 Oil     Wheat  NaturalGas    Copper    Silver     SP500
2019-01-03  0.011818  0.013814   -0.004395 -0.020762  0.010552 -0.024757
2019-01-04  0.018475  0.006326    0.033616  0.031511 -0.000700  0.034336
2019-01-05  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
2019-01-06  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
2019-01-07  0.011676 -0.000484    0.000329 -0.003960 -0.001657  0.007010
```

[65]:
```python
returns_p1_all = combined_returns.loc[PERIOD_1_START:PERIOD_1_END]
returns_p2_all = combined_returns.loc[PERIOD_2_START:]

# full correlation matrices for both periods
corr_p1_all = returns_p1_all.corr()
corr_p2_all = returns_p2_all.corr()

# (b.1) table of correlations from period 1
corr_p1_sp500 = corr_p1_all[['SP500']].drop('SP500')

print(f"\n--- (b.i) Commodity-Equity Correlations: Period 1 (ends␣
 ↪{PERIOD_1_END}) ---")
print(corr_p1_sp500)


# (b.2) a table of correlations from period 2
corr_p2_sp500 = corr_p2_all[['SP500']].drop('SP500')

print(f"\n--- (b.ii) Commodity-Equity Correlations: Period 2 (starts␣
 ↪{PERIOD_2_START}) ---")
print(corr_p2_sp500)
```

```
--- (b.i) Commodity-Equity Correlations: Period 1 (ends 2022-02-20) ---
               SP500
Oil         0.154150
Wheat       0.065429
NaturalGas  0.118617
Copper      0.305902
Silver      0.172872

--- (b.ii) Commodity-Equity Correlations: Period 2 (starts 2022-02-21) ---
               SP500
Oil         0.127969
Wheat      -0.014935
NaturalGas  0.100123
Copper      0.207982
Silver      0.218555
```

[94]:

```python
import matplotlib.pyplot as plt
import seaborn as sns

fig, axes = plt.subplots(1, 2, figsize=(10, 4), sharey=True)

# Period 1
sns.heatmap(
    corr_p1_sp500,
    ax=axes[0],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
    annot=True, fmt=".2f",
    cbar=False
)
axes[0].set_title(f"(b.i) Correlation with S&P 500\n{PERIOD_1_START}-
{PERIOD_1_END}")
axes[0].set_ylabel("Commodity")
axes[0].set_xlabel("")

# Period 2
sns.heatmap(
    corr_p2_sp500,
    ax=axes[1],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
    annot=True, fmt=".2f",
    cbar=True
)
axes[1].set_title(f"(b.ii) Correlation with S&P 500\n{PERIOD_2_START} onward")
axes[1].set_ylabel("")
axes[1].set_xlabel("")

plt.tight_layout()
plt.show()
```



- Based on the data, the extent of financialization has changed in a significant but nuanced way. Rather than a uniform increase, the results show a breakdown in financialization (a de-coupling) for the commodities most central to the geopolitical event. The correlation between the S&P 500 and Oil (0.154 to 0.128), Wheat (0.065 to -0.015), and Natural Gas (0.119 to 0.100) all decreased. This suggests that in Period 2, these commodities were no longer trading in line with general market risk appetite as represented by the S&P 500.
- The explanation for this de-coupling is that the Russian invasion of Ukraine was an over-powering, idiosyncratic supply-side shock, not a typical macroeconomic event. In Period 2, the prices of Oil, Wheat, and Natural Gas were being driven almost exclusively by their own powerful, unique fundamentals, namely, war news, sanctions, and physical supply disruptions. These factors were entirely separate from the drivers of the S&P 500 (like US corporate earnings or tech valuations). Therefore, these commodities began to trade on their own specific risk factors, breaking the risk-on/risk-off link that normally ties them to equity markets.
- The metals, Copper and Silver, tell a slightly different story. Copper's correlation, while remaining the highest, also fell (0.306 to 0.208), showing it too was partially decoupled. Silver was the only asset to show an increase in financialization (0.173 to 0.219). This is likely because as the commodity price spike caused global inflation, Silver (as a precious metal) and the S&P 500 both began reacting to the same new dominant macroeconomic factor, inflation data and the resulting interest rate hikes from central banks. This common factor re-established a link for Silver, while the raw commodities remained decoupled.

[ ]:

```

```

0.0.3 Question 1 (c)

[72]:

```
BTC_FILENAME = 'Bitcoin.csv'
PRICE_COLUMN = 'Adj Close'

# Load and process the Bitcoin data
df_btc = pd.read_csv(BTC_FILENAME)
df_btc['Date'] = pd.to_datetime(df_btc['Date'])

# same robust cleaning step we used before
df_btc[PRICE_COLUMN] = df_btc[PRICE_COLUMN].astype(str).str.replace(r'[$,]',␣
↪'', regex=True)
df_btc[PRICE_COLUMN] = pd.to_numeric(df_btc[PRICE_COLUMN], errors='coerce')

# Align to the master calendar
df_btc = df_btc.set_index('Date')

# Rename the price column
df_btc_renamed = df_btc[[PRICE_COLUMN]].rename(columns={PRICE_COLUMN: 'BTC'})
df_btc_aligned = df_btc_renamed.reindex(all_calendar_days)
```

```
# Fill missing values
df_btc_filled = df_btc_aligned.ffill()

# Backfill any NaNs at the very start
df_btc_filled = df_btc_filled.bfill()

print("--- BTC Aligned Prices (Head) ---")
print(df_btc_filled.head())
```

```
--- BTC Aligned Prices (Head) ---
                    BTC
2019-01-02  3943.409424
2019-01-03  3836.741211
2019-01-04  3857.717529
2019-01-05  3845.194580
2019-01-06  4076.632568
```

[74]:

```
# Calculate BTC 500 daily returns
btc_returns = df_btc_filled.pct_change()
combined_returns_final = combined_returns.join(btc_returns)

# Drop the first row which is all NaN
combined_returns_final = combined_returns_final.dropna(how='all')

print("\n--- Combined Returns with BTC (Head) ---")
print(combined_returns_final.head())
```

```

--- Combined Returns with BTC (Head) ---
                 Oil     Wheat  NaturalGas    Copper    Silver     SP500  \
2019-01-03  0.011818  0.013814   -0.004395 -0.020762  0.010552 -0.024757
2019-01-04  0.018475  0.006326    0.033616  0.031511 -0.000700  0.034336
2019-01-05  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
2019-01-06  0.000000  0.000000    0.000000  0.000000  0.000000  0.000000
2019-01-07  0.011676 -0.000484    0.000329 -0.003960 -0.001657  0.007010

                 BTC
2019-01-03 -0.027050
2019-01-04  0.005467
2019-01-05 -0.003246
2019-01-06  0.060189
2019-01-07 -0.012605
```

[76]:

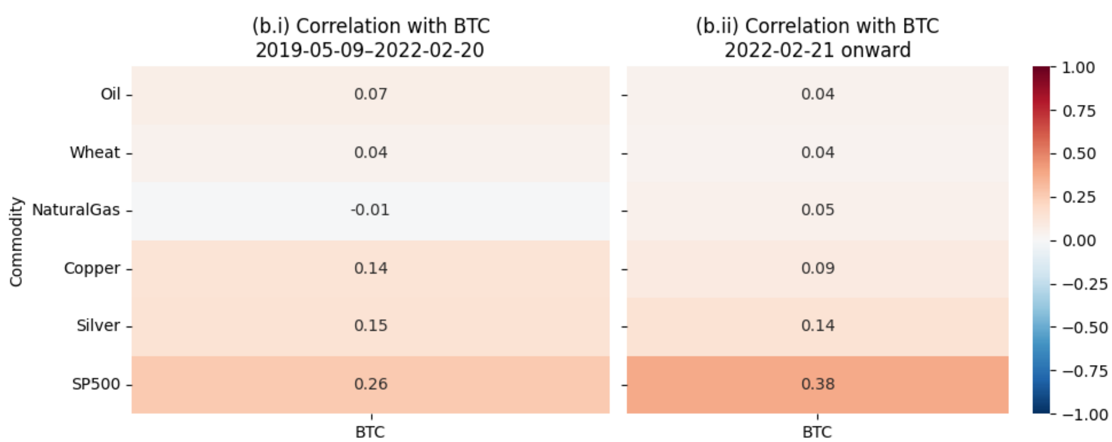
```
returns_p1_all_final = combined_returns_final.loc[PERIOD_1_START:PERIOD_1_END]
returns_p2_all_final = combined_returns_final.loc[PERIOD_2_START:]

# full correlation matrices for both periods
corr_p1_all_final = returns_p1_all_final.corr()
```

```
corr_p2_all_final = returns_p2_all_final.corr()

# (b.1) table of correlations from period 1
corr_p1_btc = corr_p1_all_final[['BTC']].drop('BTC')

print(f"\n--- (b.i) Commodity-Equity Correlations: Period 1 (ends␣
↪{PERIOD_1_END}) ---")
print(corr_p1_btc)


# (b.2) a table of correlations from period 2
corr_p2_btc = corr_p2_all_final[['BTC']].drop('BTC')

print(f"\n--- (b.ii) Commodity-Equity Correlations: Period 2 (starts␣
↪{PERIOD_2_START}) ---")
print(corr_p2_btc)
```

```
--- (b.i) Commodity-Equity Correlations: Period 1 (ends 2022-02-20) ---
                 BTC
Oil         0.068760
Wheat       0.043205
NaturalGas -0.014646
Copper      0.137085
Silver      0.145695
SP500       0.264765

--- (b.ii) Commodity-Equity Correlations: Period 2 (starts 2022-02-21) ---
                 BTC
Oil         0.042922
Wheat       0.037752
NaturalGas  0.051226
Copper      0.089849
Silver      0.141124
SP500       0.382210
```

[96]:
```
import matplotlib.pyplot as plt
import seaborn as sns

fig, axes = plt.subplots(1, 2, figsize=(10, 4), sharey=True)

# Period 1
sns.heatmap(
    corr_p1_btc,
    ax=axes[0],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
```

```
    annot=True, fmt=".2f",
    cbar=False
)
axes[0].set_title(f"(b.i) Correlation with BTC\n{PERIOD_1_START}-
{PERIOD_1_END}")
axes[0].set_ylabel("Commodity")
axes[0].set_xlabel("")

# Period 2
sns.heatmap(
    corr_p2_btc,
    ax=axes[1],
    vmin=-1, vmax=1, center=0,
    cmap="RdBu_r",
    annot=True, fmt=".2f",
    cbar=True
)
axes[1].set_title(f"(b.ii) Correlation with BTC\n{PERIOD_2_START} onward")
axes[1].set_ylabel("")
axes[1].set_xlabel("")

plt.tight_layout()
plt.show()
```



In Period 1, Bitcoin showed very low, almost random, correlations with physical commodities like Oil (0.069), Wheat (0.043), and Natural Gas (-0.015). Its strongest relationship was a moderate positive correlation with the S&P 500 (0.265), suggesting it was already more linked to financial market risk than to commodity supply and demand.

The change from Period 1 to Period 2 is the most telling observation. In Period 2, Bitcoin's correlation with almost every physical commodity decreased, moving closer to zero. However, its correlation with the S&P 500 strengthened significantly, jumping from 0.265 to 0.382. This data indicates that as the macroeconomic environment was reshaped by the inflation shock, Bitcoin did not behave like a physical commodity (which, as seen in Part (b), decoupled from the S&P 500) nor as a safe-haven currency (which would be negatively correlated).

Therefore, based on its trading behavior over the past 5 years, Bitcoin does not act like a traditional commodity or currency. Instead, it has increasingly become a speculative, risk-on financial asset. The strengthening link to the S&P 500 suggests investors have grouped Bitcoin with high-growth technology stocks, and its value is being driven by the same macroeconomic factors (like interest rate expectations and investor risk appetite) that drive the broader equity market, not by its own unique utility or supply dynamics.

[ ]:

```

```